CLAIMS

1. A method for speech synthesis, comprising:

providing a segment inventory comprising, for a plurality of speech segments, respective sequences of feature vectors, by estimating spectral envelopes of input speech signals corresponding to the speech segments in a succession of time intervals during each of the speech segments, and integrating the spectral envelopes over a plurality of window functions in a frequency domain so as to determine vector elements of the feature vectors;

receiving phonetic and prosodic information indicative of an output speech signal to be generated;

selecting the sequences of feature vectors from the inventory responsive to the phonetic and prosodic information;

processing the selected sequences of feature vectors so as to generate a concatenated output series of feature vectors;

computing a series of complex line spectra of the output signal from the series of the feature vectors; and

transforming the complex line spectra to a time domain speech signal for output.

2. A method according to claim 1, wherein providing the segment inventory comprises providing segment information comprising respective phonetic identifiers of the segments, and wherein selecting the sequences of feature vectors comprises finding the segments whose phonetic identifiers are close to the received phonetic information.

3. A method according to claim 2, wherein the segments comprise lefemes, and wherein the phonetic identifiers comprise lefeme labels.

4. A method according to claim 2, wherein the segment information further comprises one or more prosodic parameters with respect to each of the segments, and wherein selecting the sequences of feature vectors comprises finding the segments whose one or more prosodic parameters are close to the received prosodic information.

5. A method according to claim 4, wherein the one or more prosodic parameters are selected from a group of parameters consisting of a duration, an energy level and a pitch of each of the segments.

6. A method according to claim 1, wherein the feature vectors comprise auxiliary vector elements indicative of further features of the speech segments, in addition to the elements determined by integrating the spectral envelopes of the input speech signals.

7. A method according to claim 6, wherein the auxiliary vector elements comprise voicing vector elements indicative of a degree of voicing of frames of the corresponding speech segments, and wherein computing the complex line spectra comprises reconstructing the output speech signal with the degree of voicing indicated by the voicing vector elements.

8. A method according to claim 7, wherein receiving the prosodic information comprises receiving pitch values, and wherein reconstructing the output speech signal

comprises adjusting a frequency spectrum of the output speech signal responsive to the pitch values.

9. A method according to claim 1, wherein selecting the sequences of feature vectors comprises:

selecting candidate segments from the inventory;

computing a cost function for each of the candidate segments responsive to the phonetic and prosodic information and to the feature vectors of the candidate segments; and

selecting the segments so as to minimize the cost function.

10. A method according to claim 1, wherein concatenating the selected sequences of feature vectors comprises adjusting the feature vectors responsive to the prosodic information.

11. A method according to claim 10, wherein the prosodic information comprises respective durations of the segments to be incorporated in the output speech signal, and wherein adjusting the feature vectors comprises removing one or more of the feature vectors from the selected sequences so as to shorten the durations of one or more of the segments.

12. A method according to claim 10, wherein the prosodic information comprises respective durations of the segments to be incorporated in the output speech signal, and wherein adjusting the feature vectors comprises adding one or more further feature vectors to the selected sequences so as to lengthen the durations of one or more of the segments.

13. A method according to claim 10, wherein the prosodic information comprises respective energy levels of the segments to be incorporated in the output speech signal, and wherein adjusting the feature vectors comprises altering one or more of the vector elements so as to adjust the energy levels of one or more of the segments.

14. A method according to claim 1, wherein processing the selected sequences comprises adjusting the vector elements so as to provide a smooth transition between the segments in the time domain signal.

15. A method according to claim 1, wherein the vector elements comprise Mel Frequency Cepstral Coefficients of the speech segments, determined based on the integrated spectral envelopes.

16. A method for speech synthesis, comprising:

receiving an input speech signal containing a set of speech segments;

estimating spectral envelopes of the input speech signal in a succession of time intervals during each of the speech segments;

integrating the spectral envelopes over a plurality of window functions in a frequency domain so as to determine elements of feature vectors corresponding to the speech segments; and

reconstructing an output speech signal by concatenating the feature vectors corresponding to a sequence of the speech segments.

17. A method according to claim 16, wherein receiving the input speech signal comprises dividing the input speech signal into the segments and determining segment

information comprising respective phonetic identifiers of the segments, and wherein reconstructing the output speech signal comprises selecting the segments whose feature vectors are to be concatenated responsive to the segment information determined with respect to the segments.

18. A method according to claim 17, wherein dividing the input speech signal into the segments comprises dividing the signal into lefemes, and wherein the phonetic identifiers comprise lefeme labels.

19. A method according to claim 17, wherein determining the segment information further comprises finding respective segment parameters including one or more of a duration, an energy level and a pitch of each of the segments, responsive to which parameters the segments are selected for use in reconstructing the output speech signal.

20. A method according to claim 19, wherein reconstructing the output speech signal comprises modifying the feature vectors of the selected segments so as to adjust the segment parameters of the segments in the output speech signal.

21. A method according to claim 16, and comprising determining respective degrees of voicing of the speech segments, and incorporating the degrees of voicing as elements of the feature vectors for use in reconstructing the output speech signal.

22. A method according to claim 16, wherein concatenating the feature vectors comprises concatenating the vectors to form a series in a frequency domain, and

wherein reconstructing the output speech signal comprises computing a series of complex line spectra of the output signal from the series of feature vectors, and transforming the complex line spectra to a time domain signal.

23. A method according to claim 16, wherein the window functions are non-zero only within different, respective spectral windows and have variable values over their respective windows, and wherein integrating the spectral envelopes comprises calculating products of the spectral envelopes with the window functions, and calculating integrals of the products over the respective windows of the window functions.

24. A method according claim 23, and comprising applying a mathematical transformation to the integrals in order to determine the elements of the feature vectors.

25. A method according to claim 24, wherein the frequency domain comprises a Mel frequency domain, and wherein applying the mathematical transformation comprises applying log and discrete cosine transform operations in order to determine Mel Frequency Cepstral Coefficients to be used as the elements of the feature vectors.

26. A device for speech synthesis, comprising:

a memory, arranged to hold a segment inventory comprising, for a plurality of speech segments, respective sequences of feature vectors having vector elements determined by estimating spectral envelopes of input speech signals corresponding to the speech segments in a succession of time intervals during each of the speech segments, and integrating the spectral envelopes over a plurality of window functions in a frequency domain; and

a speech processor, arranged to receive phonetic and prosodic information indicative of an output speech signal to be generated, to select the sequences of feature vectors from the inventory responsive to the phonetic and prosodic information, to process the selected sequences of feature vectors so as to generate a concatenated output series of feature vectors, and to compute a series of complex line spectra of the output signal from the series of the feature vectors and transform the complex line spectra to a time domain speech signal for output.

27. A device according to claim 26, wherein the segment inventory comprises segment information comprising respective phonetic identifiers of the segments, and wherein the processor is arranged to select the sequences of feature vectors by finding the segments in the inventory whose phonetic identifiers are close to the received phonetic information.

28. A device according to claim 27, wherein the segments comprise lefemes, and wherein the phonetic identifiers comprise lefeme labels.

29. A device according to claim 27, wherein the segment information further comprises one or more prosodic parameters with respect to each of the segments, and wherein the processor is arranged to select the sequences of feature vectors by finding the segments whose one or more prosodic parameters are close to the received prosodic information.

30. A device according to claim 29, wherein the one or more prosodic parameters are selected from a group of parameters consisting of a duration, an energy level and a pitch of each of the segments.

31. A device according to claim 26, wherein the feature vectors comprise auxiliary vector elements indicative of further features of the speech segments, in addition to the elements determined by integrating the spectral envelopes of the input speech signals.

32. A device according to claim 31, wherein the auxiliary vector elements comprise voicing vector elements indicative of a degree of voicing of frames of the corresponding speech segments, and wherein the processor is arranged to reconstruct the output speech signal with the degree of voicing indicated by the voicing vector elements.

33. A device according to claim 32, wherein the prosodic information comprises pitch values, and wherein the processor is arranged to adjust a frequency spectrum of the output speech signal responsive to the pitch values.

34. A device according to claim 26, wherein the processor is arranged to select the sequences of feature vectors by selecting candidate segments from the inventory, computing a cost function for each of the candidate segments responsive to the phonetic and prosodic information and to the feature vectors of the candidate segments, and selecting the segments so as to minimize the cost function.

35. A device according to claim 26, wherein the processor is arranged to adjust the feature vectors in

the combined output series responsive to the prosodic information.

36. A device according to claim 35, wherein the prosodic information comprises respective durations of the segments to be incorporated in the output speech signal, and wherein the processor is arranged to adjust the feature vectors by removing one or more of the feature vectors from the selected sequences so as to shorten the durations of one or more of the segments.

37. A device according to claim 35, wherein the prosodic information comprises respective durations of the segments to be incorporated in the output speech signal, and wherein the processor is arranged to adjust the feature vectors by adding one or more further feature vectors to the selected sequences so as to lengthen the durations of one or more of the segments.

38. A device according to claim 35, wherein the prosodic information comprises respective energy levels of the segments to be incorporated in the output speech signal, and wherein the processor is arranged to adjust the energy levels of one or more of the segments by altering one or more of the vector elements.

39. A device according to claim 26, wherein the processor is arranged to adjust the vector elements so as to provide a smooth transition between the segments in the time domain signal.

40. A device according to claim 26, wherein the vector elements comprise Mel Frequency Cepstral Coefficients of the speech segments, determined based on the integrated spectral envelopes.

41. A device for speech synthesis, comprising:

a memory, arranged to hold a segment inventory determined by processing an input speech signal containing a set of speech segments so as to estimate spectral envelopes of the input speech signal in a succession of time intervals during each of the speech segments, and integrating the spectral envelopes over a plurality of window functions in a frequency domain so as to determine elements of feature vectors corresponding to the speech segments; and

a speech processor, arranged to reconstruct an output speech signal by concatenating the feature vectors corresponding to a sequence of the speech segments.

42. A device according to claim 41, wherein the input speech signal is processed by dividing the input speech signal into the segments and determining segment information comprising respective phonetic identifiers of the segments, and wherein the processor is arranged to reconstruct the output speech signal by selecting the segments whose feature vectors are to be concatenated responsive to the segment information determined with respect to the segments.

43. A device according to claim 42, wherein the input speech signal is divided into lefemes, and the phonetic identifiers comprise lefeme labels.

44. A device according to claim 42, wherein the segment information further comprises respective segment parameters including one or more of a duration, an energy level and a pitch of each of the segments, responsive to which parameters the segments are selected by the

processor for use in reconstructing the output speech signal.

45. A device according to claim 44, wherein the processor is arranged to modify the feature vectors of the selected segments so as to adjust the segment parameters of the segments in the output speech signal.

46. A device according to claim 41, wherein the feature vectors comprise respective degrees of voicing of the speech segments, for use by the processor in reconstructing the output speech signal.

47. A device according to claim 41, wherein the processor is arranged to concatenate the feature vectors to form a series in a frequency domain, and to reconstruct the output speech signal by computing a series of complex line spectra of the output signal from the series of feature vectors, and transforming the complex line spectra to a time domain signal.

48. A device according to claim 14, wherein the window functions are non-zero only within different, respective spectral windows and have variable values over their respective windows, and wherein the feature vector elements are determined by calculating products of the spectral envelopes with the window functions, and calculating integrals of the products over the respective windows of the window functions.

49. A device according to claim 48, wherein a mathematical transformation is applied to the integrals in order to determine the elements of the feature vectors.

50. A device according to claim 48, wherein the frequency domain comprises a Mel frequency domain, and

wherein the mathematical transformation comprises log and discrete cosine transform operations, which are applied so as to determine Mel Frequency Cepstral Coefficients to be used as the elements of the feature vectors.

51. A computer software product, comprising a computer-readable medium in which program instructions are stored, which instructions, when read by a computer, cause the computer to access a segment inventory comprising, for a plurality of speech segments, respective sequences of feature vectors having vector elements determined by estimating spectral envelopes of input speech signals corresponding to the speech segments in a succession of time intervals during each of the speech segments, and integrating the spectral envelopes over a plurality of window functions in a frequency domain, and in response to phonetic and prosodic information indicative of an output speech signal to be generated, cause the computer to select the sequences of feature vectors from the inventory responsive to the phonetic and prosodic information, to process the selected sequences of feature vectors so as to generate a concatenated output series of feature vectors, and to compute a series of complex line spectra of the output signal from the series of the feature vectors and transform the complex line spectra to a time domain speech signal for output.

52. A product according to claim 51, wherein the segment inventory comprises segment information comprising respective phonetic identifiers of the segments, and wherein the instructions cause the computer to select the sequences of feature vectors by finding the segments in

the inventory whose phonetic identifiers are close to the received phonetic information.

53. A product according to claim 52, wherein the segments comprise lefemes, and wherein the phonetic identifiers comprise lefeme labels.

54. A product according to claim 52, wherein the segment information further comprises one or more prosodic parameters with respect to each of the segments, and wherein the instructions cause the computer to select the sequences of feature vectors by finding the segments whose one or more prosodic parameters are close to the received prosodic information.

55. A product according to claim 54, wherein the one or more prosodic parameters are selected from a group of parameters consisting of a duration, an energy level and a pitch of each of the segments.

56. A product according to claim 54, wherein the feature vectors comprise auxiliary vector elements indicative of further features of the speech segments, in addition to the elements determined by integrating the spectral envelopes of the input speech signals.

57. A product according to claim 56, wherein the auxiliary vector elements comprise voicing vector elements indicative of a degree of voicing of frames of the corresponding speech segments, and wherein the instructions cause the computer to reconstruct the output speech signal with the degree of voicing indicated by the voicing vector elements.

58. A product according to claim 57, wherein the prosodic information comprises pitch values, and wherein

the instructions cause the computer to adjust a frequency spectrum of the output speech signal responsive to the pitch values.

59. A product according to claim 51, wherein the instructions cause the computer to select the sequences of feature vectors by selecting candidate segments from the inventory, computing a cost function for each of the candidate segments responsive to the phonetic and prosodic information and to the feature vectors of the candidate segments, and selecting the segments so as to minimize the cost function.

60. A product according to claim 51, wherein the instructions cause the computer to adjust the feature vectors in the combined output series responsive to the prosodic information.

61. A product according to claim 60, wherein the prosodic information comprises respective durations of the segments to be incorporated in the output speech signal, and wherein the instructions cause the computer to adjust the feature vectors by removing one or more of the feature vectors from the selected sequences so as to shorten the durations of one or more of the segments.

62. A product according to claim 60, wherein the prosodic information comprises respective durations of the segments to be incorporated in the output speech signal, and wherein the instructions cause the computer to adjust the feature vectors by adding one or more further feature vectors to the selected sequences so as to lengthen the durations of one or more of the segments.

63. A product according to claim 60, wherein the prosodic information comprises respective energy levels of the segments to be incorporated in the output speech signal, and wherein the instructions cause the computer to adjust the energy levels of one or more of the segments by altering one or more of the vector elements.

64. A product according to claim 51, wherein the instructions cause the computer to adjust the vector elements so as to provide a smooth transition between the segments in the time domain signal.

65. A product according to claim 51, wherein the vector elements comprise Mel Frequency Cepstral Coefficients of the speech segments, determined based on the integrated spectral envelopes.

66. A computer software product, comprising a computer-readable medium in which a segment inventory is stored, the inventory having been determined by processing an input speech signal containing a set of speech segments so as to estimate spectral envelopes of the input speech signal in a succession of time intervals during each of the speech segments, and integrating the spectral envelopes over a plurality of window functions in a frequency domain so as to determine elements of feature vectors corresponding to the speech segments, so that a speech processor can reconstruct an output speech signal by concatenating the feature vectors corresponding to a sequence of the speech segments.

67. A product according to claim 66, wherein the input speech signal is processed by dividing the input speech signal into the segments and determining segment

information comprising respective phonetic identifiers of the segments, and wherein to reconstruct the output speech signal, the processor selects the segments whose feature vectors are to be concatenated responsive to the segment information determined with respect to the segments.

68. A product according to claim 66, wherein the input speech signal is divided into lefemes, and the phonetic identifiers comprise lefeme labels.

69. A product according to claim 66, wherein the segment information further comprises respective segment parameters including one or more of a duration, an energy level and a pitch of each of the segments, responsive to which parameters the segments are selected by the computer for use in reconstructing the output speech signal.

70. A product according to claim 69, wherein to reconstruct the output speech signal, the instructions cause the computer to modify the feature vectors of the selected segments so as to adjust the durations and energy levels of the segments in the output speech signal.

71. A product according to claim 66, wherein the feature vectors comprise respective degrees of voicing of the speech segments, for use by the computer in reconstructing the output speech signal.

72. A product according to claim 66, wherein to reconstruct the output speech signal, the instructions cause the computer to concatenate the feature vectors to form a series in a frequency domain, to compute as series

of complex line spectra of the output signal from the series of feature vectors, and to transform the complex line spectra to a time domain signal.

73. A product according to claim 66, wherein the window functions are non-zero only within different, respective spectral windows and have variable values over their respective windows, and wherein the feature vector elements are determined by calculating products of the spectral envelopes with the window functions, and calculating integrals of the products over the respective windows of the window functions.

74. A product according claim 73, wherein a mathematical transformation is applied to the integrals in order to determine the elements of the feature vectors.

75. A product according to claim 74, wherein the frequency domain comprises a Mel frequency domain, and wherein the mathematical transformation comprises log and discrete cosine transform operations, which are applied so as to determine Mel Frequency Cepstral Coefficients to be used as the elements of the feature vectors.